

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-204511

(43)Date of publication of application : 05.08.1997

(51)Int.Cl.

G06T 1/00  
G06F 17/30

(21)Application number : 08-010170

(71)Applicant : SHARP CORP

(22)Date of filing : 24.01.1996

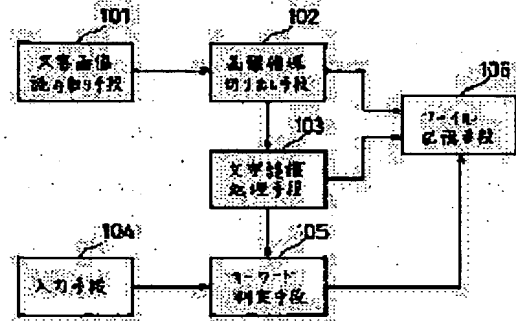
(72)Inventor : ISHIHARA TOMOYUKI

## (54) FILING DEVICE

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To file only the article that includes the key word set by a user in its header among the articles of newspapers, magazines, etc.

**SOLUTION:** This device includes an image sensor 101, a keyboard 104 which inputs a key word, and a CPU. The CPU segments (102) a header and the article of the text corresponding to the header out of the document images of article pages of newspapers, magazines, etc., which are read by the sensor 101. Then the CPU performs the character recognition processing 103 for the segmented header to acquire its character codes and decides (105) whether the key word is included in the character codes of the header. Then the character codes of the header including the key word and the document images of the article pages corresponding to the header are filed in a storage 106 with relation secured between them.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁(JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 9 - 2 0 4 5 1 1

(43) 公開日 平成9年(1997)8月5日

(51) Int. Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 T	1/00		G 0 6 F 15/62	3 3 0 D
G 0 6 F	17/30		15/40	3 7 0 B
			15/401	3 3 0 Z

審査請求 未請求 請求項の数 4

O L

(全 9 頁)

(21) 出願番号 特願平8-10170

(22) 出願日 平成8年(1996)1月24日

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 石原 朋幸

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

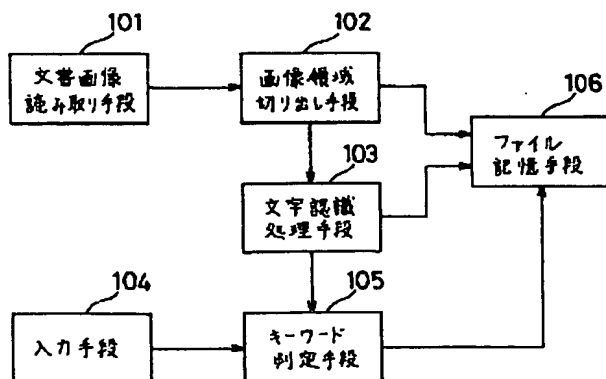
(74) 代理人 弁理士 野河 信太郎

(54) 【発明の名称】 ファイリング装置

(57) 【要約】

【課題】 新聞・雑誌等の記事の内、使用者の設定したキーワードが見出しに含まれている記事だけファイリングする。

【解決手段】 イメージセンサと、キーワードを入力するキーボードと、CPUを備え、CPUにより、イメージセンサで読み取った新聞・雑誌等の記事面の文書画像から見出しとその見出しに対応する本文の記事を切り出し、切り出した見出しについて文字認識処理を行い見出しの文字コードを得、見出しの文字コードにキーワードが含まれているか否かを判定し、キーワードが含まれている見出しの文字コードとその見出しに対応する本文の記事の文書画像とを関連づけて記憶装置にファイリングする。



## 【特許請求の範囲】

【請求項 1】 新聞・雑誌等の記事面を文書画像として光学的に読み取る文書画像読み取り手段と、

読み取られた文書画像から見出しを切り出すとともにその見出しに対応する本文の記事を切り出す画像領域切り出し手段と、

切り出された見出しの文書画像について文字認識処理を行い見出しの文字コードを得る文字認識処理手段と、

文字列からなるキーワードを入力する入力手段と、

見出しの文字コードにキーワードが含まれているか否かを判定するキーワード判定手段と、

キーワードが含まれている見出しの文字コードとその見出しに対応する本文の記事の文書画像とを関連づけて記憶するファイル記憶手段を備えてなるファイリング装置。

【請求項 2】 入力手段が複数のキーワードを入力することが可能であり、キーワード判定手段が、見出しの文字コードに複数のキーワードの内のいずれかが含まれているか否かを判定することを特徴とする請求項 1 記載のファイリング装置。

【請求項 3】 文書画像読み取り手段によって読み取られた文書画像から日付表示領域を切り出す日付表示領域切り出し手段と、

切り出された日付表示領域の文書画像について文字認識処理を行い日付の文字コードを得る日付認識処理手段をさらに備え、

ファイル記憶手段が、キーワードが含まれている見出しの文字コードとその見出しに対応する本文の記事の文書画像とを日付の文字コードに関連づけて記憶することを特徴とする請求項 1 記載のファイリング装置。

【請求項 4】 文書画像読み取り手段によって読み取られた文書画像から日付表示領域を切り出す日付表示領域切り出し手段と、

切り出された日付表示領域の文書画像について文字認識処理を行い日付の文字コードを得る日付認識処理手段と、

日付の文字コードを文字画像化して本文の記事の文書画像に追加する追加手段をさらに備え、

ファイル記憶手段が、キーワードが含まれている見出しの文字コードとその見出しに対応する本文の記事の文書画像とを日付の文字コードに関連づけて記憶することを特徴とする請求項 1 記載のファイリング装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 この発明は、新聞・雑誌等の記事を文書画像として読み取ってファイリングするファイリング装置に関する。

## 【0002】

【従来の技術】 従来、この種のファイリング装置としては、特開平 3-14184 号公報に記載のような文書画

像再配置ファイリング装置が知られている。これにおいては、新聞・雑誌等の記事を文書画像として読み取った後、見出しの領域と本文の領域とを自動的に判別し、各々の文書画像の領域を切り出し、これらの画像領域の各々に対して文字認識処理を施して文字コードデータを得、見出しのデータと本文のデータとを別々に管理する形でファイリングしておき、呼び出す際は文字コードデータに含まれる文字列を入力指定する方法により見出しのデータ、本文のデータをそれぞれ別々に検索し、目的の見出しのデータと本文のデータを再配置して表示するようにしている。

## 【0003】

【発明が解決しようとする課題】 しかしながら、このような従来のファイリング装置においては、使用者が必要としない記事も含めた全ての記事の見出し領域、本文領域の文書画像に対して文字認識処理を行なうようにしているため、ファイリングの際、かなりの処理時間を必要とする。そして、使用者が必要としない記事も含めた全ての記事のデータをファイリングするため、膨大な容量の記憶装置が必要となる。

【0004】 また、本文と見出しとを関連づけてファイリングしていないため、ファイリングされている一つの記事を画面表示もしくは印刷する際には、見出しに対応する本文を使用者が判断し、見出し及び本文をそれぞれ呼び出した上で画面表示もしくは印刷する必要があり、記事の呼び出しに手数を必要とする。

【0005】 さらに、記事のデータをファイリングする際に、新聞等の発行日の情報をファイリングしていないため、記事のデータがいつ出版されたものであるのかを知ることができないという問題があった。

【0006】 この発明は、このような事情を考慮してなされたもので、新聞・雑誌等の記事の内、使用者の設定したキーワードが見出しに含まれている記事についてだけ、見出しの文字コードとその見出しに対応する本文の記事の文書画像とを関連づけてファイリングするようにしたファイリング装置を提供するものである。

## 【0007】

【課題を解決するための手段】 図 1 はこの発明の構成を示すブロック図である。この図に示すように、この発明は、新聞・雑誌等の記事面を文書画像として光学的に読み取る文書画像読み取り手段 101 と、読み取られた文書画像から見出しを切り出すとともにその見出しに対応する本文の記事を切り出す画像領域切り出し手段 102 と、切り出された見出しの文書画像について文字認識処理を行い見出しの文字コードを得る文字認識処理手段 103 と、文字列からなるキーワードを入力する入力手段 104 と、見出しの文字コードにキーワードが含まれているか否かを判定するキーワード判定手段 105 と、キーワードが含まれている見出しの文字コードとその見出しに対応する本文の記事の文書画像とを関連づけて記憶

するファイル記憶手段106を備えてなるファイリング装置である。

【0008】この発明において、文書画像読み取り手段101としては、新聞・雑誌等の記事面を文書画像として光学的に読み取ることが可能なものであればよく、光学式のイメージスキャナを用いることが望ましいが、C Dカメラのような装置を用いることもできる。

【0009】画像領域切り出し手段102、文字認識処理手段103及びキーワード判定手段105としては、CPU、ROM、RAM等を含むマイクロコンピュータ 10を用いるのが便利である。

【0010】入力手段104としては、文字列を入力することが可能なキーボードを用いることが好ましいが、ペン、マウス、もしくはトラックボールのような各種のポインティングデバイスを用いることもできる。ファイル記憶手段106としては、磁気ディスク装置やフロッピーディスク装置のような各種の記憶装置を用いることができる。

【0011】この発明によれば、新聞・雑誌等の記事面が文書画像読み取り手段101によって読み取られると、読み取られた新聞・雑誌等の記事面の文書画像から、画像領域切り出し手段102によって見出しとその見出しに対応する本文の記事が切り出され、文字認識処理手段103によって、切り出された見出しの文書画像が文字認識されて見出しの文字コードが得られる。 20

【0012】そして、キーワード判定手段105によって、見出しの文字コードにキーワードが含まれているか否かが判定され、ファイル記憶手段106により、キーワードが含まれた見出しの文字コードとその見出しに対応する本文の記事の文書画像とが関連づけられて記憶される。したがって、キーワードが見出しに含まれている記事だけを選択してファイリングすることができる。 30

【0013】上記構成においては、入力手段104を複数のキーワードが入力可能な構成とし、キーワード判定手段105を、見出しの文字コードに複数のキーワードの内のいずれかが含まれているか否かを判定するように構成することが好ましい。

【0014】このように構成した場合には、見出しの文字コードに複数のキーワードのうちいずれかが含まれていれば、これらのキーワードの含まれた見出しに対応する本文の記事の文書画像が、見出しの文字コードに関連づけられて記憶される。したがって、複数のキーワードの内のいずれかが見出しに含まれている記事だけを選択してファイリングすることができる。 40

【0015】また、上記構成においては、文書画像読み取り手段101によって読み取られた文書画像から日付表示領域を切り出す日付表示領域切り出し手段と、切り出された日付表示領域の文書画像について文字認識処理を行い日付の文字コードを得る日付認識処理手段とをさらに設けた構成とし、ファイル記憶手段106を、キー 50

ワードが含まれた見出しの文字コードとその見出しに対応する本文の記事の文書画像とを日付の文字コードに関連づけて記憶するように構成することが好ましい。

【0016】このように構成した場合には、新聞等の日付表示領域から得られた日付の文字コードが、見出しの文字コードとその見出しに対応する本文の記事の文書画像とに関連づけられて記憶される。したがって、日付の文字コードで記事の発行日を知ることができる。また、記事を日付で検索することが可能となる。

【0017】上記構成においては、日付の文字コードを文字画像化して本文の記事の文書画像に追加する追加手段をさらに設けた構成とすることがより好ましい。

【0018】このように構成した場合には、新聞等の日付表示領域から得られた日付の文字コードが文字画像化されて本文の記事の文書画像に追加された上で、日付の文字コードが、見出しの文字コードとその見出しに対応する本文の記事の文書画像とに関連づけられて記憶される。したがって、本文の記事の文書画像を表示あるいは印刷したときに日付も同時に表示あるいは印刷することができる。 50

【0019】

【発明の実施の形態】以下、図面に示す実施例に基づいてこの発明を詳述する。なお、これによってこの発明が限定されるものではない。

【0020】図2はこの発明を日本語ワードプロセッサに適用した一実施例の構成を示すブロック図である。本日本語ワードプロセッサは、キーボード、CPU、メモリ、表示装置、印字装置等の公知の基本的な装置から構成されている。したがって、以下においては、それらの基本的な装置については説明を省略する。また、仮名漢字変換機能のような日本語ワードプロセッサとしての基本機能についてはすでに備えているものとして説明を行う。

【0021】本日本語ワードプロセッサの内、ファイリング装置として機能する部分だけを説明すれば、本ファイリング装置は、キーワード登録部1と、文書画像読み取り部2と、記事領域切り出し部3と、日付表示領域切り出し部4と、見出し領域切り出し部5と、文字認識処理部6と、日付情報抽出部7と、日付追加部8と、キーワード判定部9と、データファイル記憶部10から構成されている。

【0022】上記において、キーワード登録部1は、文字列からなるキーワードを入力することが可能なキーボードと、入力されたキーワードを記憶可能なRAMから構成される。文書画像読み取り部2は、光学式のイメージセンサーから構成される。記事領域切り出し部3、日付表示領域切り出し部4、見出し領域切り出し部5、文字認識処理部6、日付情報抽出部7、日付追加部8、及びキーワード判定部9は、CPUから構成される。データファイル記憶部10は、フロッピーディスク装置、磁

気ディスク装置等の記憶装置から構成される。

【0023】キーワード登録部1は、キーボードから入力された必要な記事の見出しに含まれるキーワードをRAMに登録する。文書画像読み取り部2は、新聞・雑誌などの紙面全体の文書画像を光学的に読み取る。記事領域切り出し部3は、各記事の画像領域を切り出す。日付表示領域切り出し部4は、新聞紙面の上欄にある日付表示領域を切り出す。見出し領域切り出し部5は、各記事の見出しの領域を切り出す。

【0024】文字認識処理部6は、日付表示領域切り出し部4で切り出された文書画像と見出し領域切り出し部5で切り出された文書画像に対して文字認識処理を施して日付表示領域の文字コードデータと見出し領域の文字コードデータを得る。日付情報抽出部7は、文字認識処理部6により文字認識処理して得られた日付表示領域の文字コードデータより日付の情報を抽出する。日付追加部8は、記事領域切り出し部3で切り出された記事の画像に、日付情報抽出部7により抽出された日付情報を文字化した画像を追加する。キーワード判定部9は、文字認識処理部6により文字認識処理して得られた見出しの文字コードデータにキーワード登録部2で登録されたキーワードが含まれているか否かを判定する。

【0025】データファイル記憶部10は、キーワード判定部9によって見出しにキーワードが含まれていると判定された場合、その見出しの文字コードデータと、日付情報抽出部7で得られた日付の情報と、その見出しに含まれているキーワードと、その見出しに対応し日付追加部8により日付の文字画像が追加された記事の文書画像データ（記事画像データ）とを、それぞれ関連づけて記憶装置にファイリングする。

【0026】新聞紙面から必要な記事のファイリングを行う場合の具体的な例を、図3を用いて以下に説明する。ファイリングしたい記事の見出しに含まれるキーワードとして、例えば「AA」と「BB」の2つが使用者によって設定され、キーワード登録部1によってすでに登録されているものとする。

【0027】まず、使用者の操作によって、文書画像読み取り部2は、図3に示す新聞紙面全体の文書画像を読み取る。その後、この文書画像より新聞紙面の上欄にある日付表示領域11を切り出し、この日付表示領域11の画像に対して文字認識処理を行い文字コードデータを得る。

【0028】この切り出しの具体的な処理においては、まず、紙面全体の文書画像の上端より一定距離以内の領域内の黒画素を検出して、一定長さ以上に黒画素が水平方向に連続している場合、これを紙面上部の横罫線と認識する。次に、この横罫線より一定距離以内上側の文書画像領域を切り出し、この領域内で黒画素の連結領域を検出すると、これを文字とみなして文字認識処理を施し、文字コードデータを得る。

【0029】この文字コードデータより年、月、日、曜日の文字が検出されるので、これらの文字の前の数字及び文字を認識する事により、図3の例であれば1995年9月11日の月曜日に発行されたという情報を得、これをメモリに記憶する。

【0030】次に、紙面全体の文書画像より記事12、記事13、記事14の各領域と、それらの各領域に対応する見出し15、見出し16、見出し17の各領域とを切り出す。

【0031】具体的な方法としては、例えば紙面全体の文書画像について一定値以内の黒画素の連結を検索し、これを内包できる最小限の大きさの正方形領域を文字矩形とする。見出しの文字は一般に本文の文字より大きいので、あらかじめ規定する閾値より大きな文字矩形を検出するとこれを見出し文字であると認識し、この文字矩形と一定の誤差の範囲で同じ大きさの文字矩形が等間隔で並んでいることを検出すると、この文字矩形の並びの全体を含む領域を見出し領域であると判断して、この領域の画像データを切り出す。

【0032】次に、紙面全体の文書画像について水平方向および垂直方向に一定長さ以上の黒画素の連続を検出した場合、これを罫線であると認識し紙面を縦罫線と横罫線に囲まれたいくつかの区画に分ける。各区画をその内部にある見出し領域の指し示す本文の記事の領域であると判断して、紙面より各本文の記事の領域を切り出す。そして、見出し15、見出し16、見出し17に対して、それぞれ見出し番号1、見出し番号2、見出し番号3を付す。

【0033】次に、見出し番号1である見出し15の文書画像に対して文字認識処理を施し、文字コードデータを得る。この見出し15の文字コードデータに対して先に設定したキーワードが含まれているか否かを判定する。

【0034】具体的には、見出しの文字コードデータ列内に先に登録しておいたキーワードの文字コードデータ列と一致する文字コードデータ列がある場合には、見出しにキーワードが含まれていると判断する。これにより、本例においては、見出し番号1の見出し15に「AA」というキーワードが含まれていることがわかる。

【0035】この場合に、見出し15にキーワードが含まれているので、見出し15に対応する記事12の本文の文書画像データに、先に得た日付の情報を文字化した画像を追加した後、見出しの文字コードデータと、先に得た日付情報の数値データと、見出しに含まれているキーワードと、記事画像データとを関連づけて記憶装置にファイリングする。

【0036】ファイリングされたデータの構造は、例えば図4に示す様になる。このように、各データは、ファイル情報21、見出しの文字コードデータ22、日付データ23、キーワード24、記事画像データ25、ファ

イル終了コード26の順でファイルされる。

【0037】ファイル情報21はファイルの大きさなど一般的なファイル情報である。見出しの文字コードデータ22は記事画像データ25に対応する見出しの文字コードデータである。日付データ23は日付の数値データであり、図3の例では“19950911”という数値データである。キーワード24は見出しに含まれていたキーワードの文字コードデータであり、図3の例では“AA”という文字コードデータである。記事画像データ25は当該記事の領域の文書画像データである。ファイル終了コード26はファイルの最後を示すコードである。

【0038】このようにしてファイリングされた記事の画像は図5に示すように表示される。この表示例に示すように、記事画像データには日付情報が追加されているので、記事画像の表示時には、追加された日付情報の文字画像31が記事画像とともに表示される。

【0039】以上で見出し番号1の記事についてのファイリングを終了し、次に見出し番号2の記事についてのファイリングに移行する。見出し番号2の記事についてのファイリングでは、まず、見出し16の領域の文書画像データに対して文字認識処理を施して文字コードデータを得る。このデータに対して先に設定したキーワードが含まれているか否かを判定すると、この見出しにはキーワードは含まれていない。したがって、この見出し番号2の記事についてのファイリングは行わない。

【0040】ここまで処理した見出し番号は2であり、最後の見出しではないので次に見出し番号3の記事についてのファイリングに移行する。見出し番号3の記事についてのファイリングでは、まず、見出し17の領域の文書画像データに対して文字認識処理を施して文字コードデータを得る。このデータに対して先に設定したキーワードが含まれるか否かを判定すると、「BB」というキーワードが含まれていることがわかる。

【0041】したがって、先に切り分けられている、この見出し17に対応する記事14の領域の文書画像データに先に得た日付の情報を文字化した画像を追加した後、見出しの文字コードデータと、先に得た日付情報の数値データと、見出しに含まれているキーワードと、記事画像データとを関連づけて記憶装置にファイリングする。

【0042】ここまで処理した見出しの番号は3であり、最後の見出しである。したがって、この紙面に対する処理は終了する。次の紙面について引続き検索、ファイリング処理を行うか否かを使用者のオペレーションにより選択し、引続き処理を行う場合は次の紙面の日付表示領域の切り出しより処理を開始する。

【0043】次に、このような実施例の動作を図6及び図7に示すフローチャートに従い説明する。図6はキーワードを登録する場合の処理内容を示すフローチャート

である。キーワードの登録においては、まず、使用者は、ファイリングしたい記事の見出しに含まれる文字列をキーワードとして登録する(ステップS1)。複数のキーワードを登録したい場合は、必要なだけ登録する(ステップS2)。

【0044】具体的には、使用者がキーボードからキーワードとして登録したい文字列を入力すると、この文字列に対応した文字コードデータがRAMに記憶される。RAMにはこのように複数のメモリ領域を設定しておく。

【0045】図7はファイリングの処理内容を示すフローチャートである。ファイリングの処理においては、まず、文書画像読み取り部2によって紙面全体の文書画像を読み取り(ステップS3)、新聞の最上欄の日付が表示された画像領域を切り出して文字認識処理を行い、日付情報を得る(ステップS4)。すなわち、文字認識した文字コードデータ内で年、月、日の文字を検索して、これらの文字の前の数字により発行年月日の情報を得、これを年、月、日の数値データに変換し、専用のメモリに記憶しておく。

【0046】次に、紙面全体の画像より、各見出しの領域と、各見出しが指し示す記事の領域とを切り出す(ステップS5)。次に、切り出された各記事の見出しの領域に対して、それぞれ1番から順に番号付けをする(ステップS6)。

【0047】次に、1番目に番号付けされた見出し番号1の領域の文書画像データに対して文字認識処理を行い、見出しの文字コードデータを得(ステップS7)、見出しの文字コードデータに先に登録したキーワードのうち何れかが含まれているか否かを判定する(ステップS8)。

【0048】ここで、見出しの文字コードデータに何れのキーワードも含まれていない場合には、その見出しが指し示す記事はファイリングしないものとして、ステップS11に進む。一方、キーワードが含まれているときには、切り出された記事の領域の文書画像データに、日付の情報を文字化した画像を追加する(ステップS9)。

【0049】そして、日付情報の文字画像が追加された記事の文書画像データに、見出しの文字コードデータと、日付情報の数値データと、見出しに含まれているキーワードとを、各々関連づけて記憶装置にファイリングする(ステップS10)。

【0050】これは、例えば記事の文書画像データに、データヘッダとしてこの記事を指し示す見出しの文字コードデータ、日付の情報を示すデータ、見出しに含まれているキーワードの文字コードデータを付加して、ひとつのファイルとして記憶装置にファイリングする。

【0051】記事の文書画像データを、見出し、日付、キーワードと各々関連づけてファイリングした後は、

見出し番号からその見出しが最後の見出しであるか否かを判定する（ステップ S11）。

【0052】ここで、見出しが最後の見出しでなければ、ステップ S7へ戻って、次の見出しの領域の文書画像データに対する文字認識処理へ移る。見出しが最後の見出しであれば、次の紙面について、引続き検索、ファイリング処理を行うか否かを使用者が判断する（ステップ S12）。この使用者の判断において、次の紙面について処理を続ける場合は、ステップ S3に戻って、次の紙面に対する日付表示領域の切り出しから処理を開始する。現在の紙面で処理を完了する場合は、動作を終了する。

【0053】このようにして、新聞・雑誌等の記事の内、使用者の設定したキーワードが“見出し”に含まれている記事についてだけ、見出しの文字コードデータと、出版日の情報と、キーワードと、見出しに対応する本文の画像データとを関連づけてファイリングする。

【0054】すなわち、はじめに使用者が必要な記事の見出しに含まれるキーワードを登録しておいて、新聞・雑誌の記事のファイリングを開始すると、文書画像読み取り部 2により紙面全体を読み取った後、日付表示領域、各見出しの領域、各記事の領域を切り出す。そして、日付表示領域の文書画像に対して文字認識処理を施して日付情報を得、各見出しの領域の文書画像データに対して文字認識処理を施して見出しの文字コードデータを得、この見出しの文字コードデータにあらかじめ登録しておいたキーワードが含まれているか否かを判定して、含まれている場合には、この見出しに対応する記事の領域の文書画像データに先に得た日付情報を文字画像化して追加する。そして、この日付情報を追加した記事画像データに、見出しの文字コードデータと、先に得た日付情報と、見出しの文字コードデータに含まれていたキーワードの文字コードデータとを関連づけてファイリングする。

【0055】したがって、必要な記事を手作業で 1 つずつ切り抜いてファイリングするような時間と手間をかけずに記事のファイリングができる。また、このようなフ

ファイリング方式であれば、紙面上の全ての記事をファイリングするシステムのように不必要な記事をファイリングしないので、無駄な処理時間や無駄な記憶容量を必要としない。

【0056】さらに、ファイリングされた記事のデータと呼び出す際に、見出し文やファイリングされた際のキーワードや発行日付などを手掛かりに検索する事が容易であるし、逆に呼び出された記事データの発行日付や記事がファイリングされた際のキーワードを知る事が容易である。

【0057】

【発明の効果】この発明によれば、キーワードが見出しに含まれている記事だけを選択してファイリングすることができる。

【図面の簡単な説明】

【図 1】本発明の構成を示すブロック図である。

【図 2】本発明を日本語ワードプロセッサに適用した一実施例の構成を示すブロック図である。

【図 3】本発明により処理される新聞紙面の例を示す説明図である。

【図 4】本発明によりファイリングされたデータ構造の例を示す説明図である。

【図 5】本発明によりファイリングされた記事の表示例を示す説明図である。

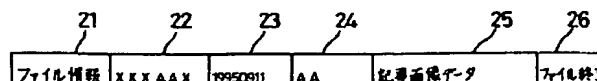
【図 6】実施例の動作を示すフローチャートである。

【図 7】実施例の動作を示すフローチャートである。

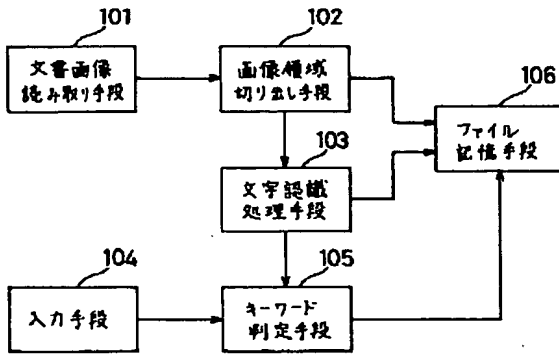
【符号の説明】

- 1 キーワード登録部
- 2 文書画像読み取り部
- 3 記事領域切り出し部
- 4 日付表示領域切り出し部
- 5 見出し領域切り出し部
- 6 文字認識処理部
- 7 日付情報抽出部
- 8 日付追加部
- 9 キーワード判定部
- 10 データファイル記憶部

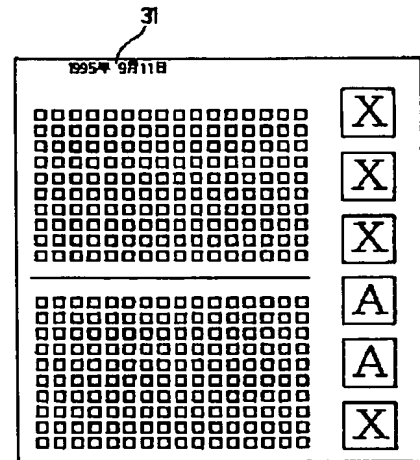
【図 4】



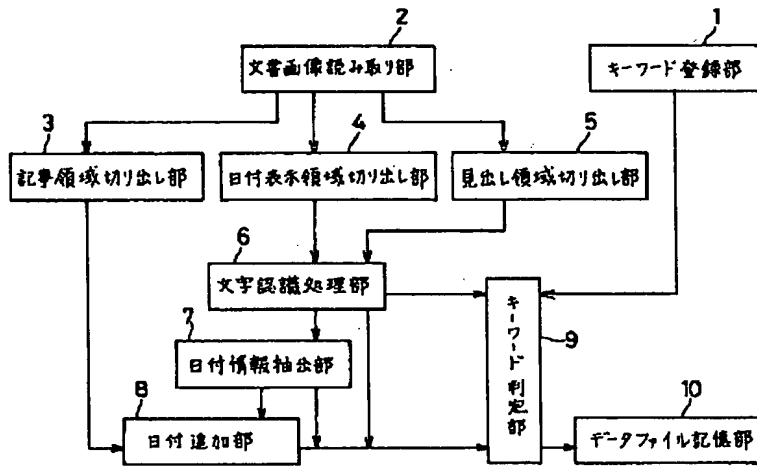
【図1】



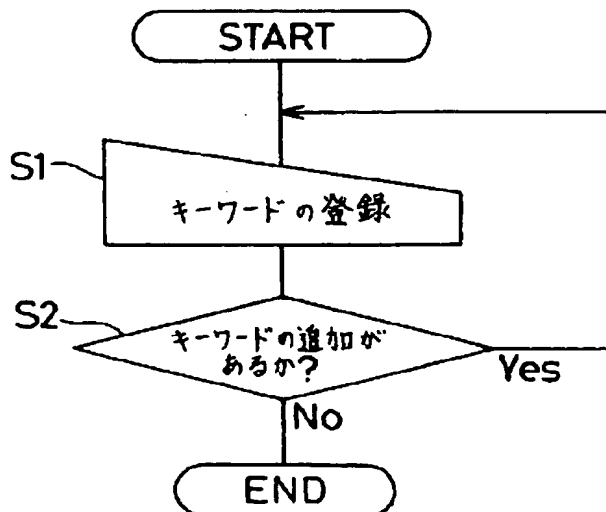
【図5】



【図2】

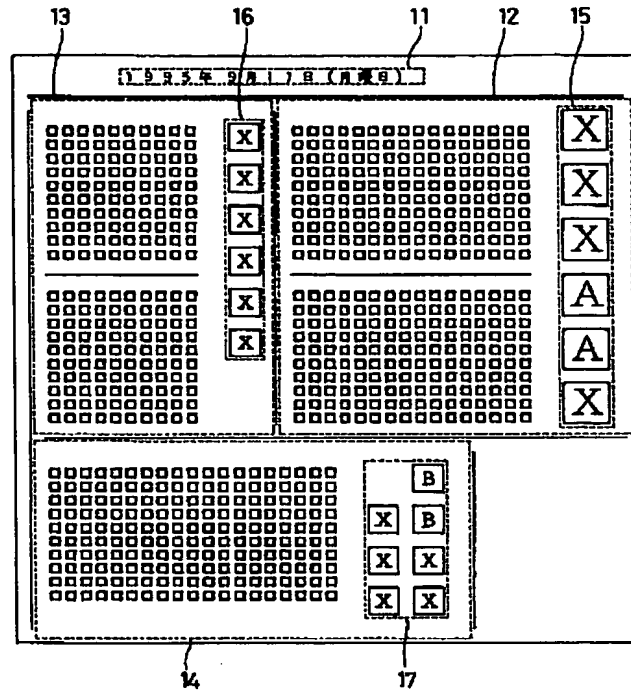


【図6】





【図3】



【図7】

